A research information system as a research planning and evaluation tool recent developments in Denmark

Price, Adrian

Published in:
International Conference on Current Research Information Systems

Publication date:
2008

Document version
Publisher's PDF, also known as Version of record

A research information system as a research planning and evaluation tool: recent developments in Denmark

Adrian Price, The Faculty of Life Sciences Library, Faculty of Life Sciences, University of Copenhagen

Summary

CRIS (Current Research Information System) developments at universities in Denmark have changed dramatically over the last 2-3 years. Most universities have adopted the same software for the registration and documentation of the research activities of researchers. This means that formats are standardized which means that their research activities can be compared and benchmarked. At the same time, the system has been developed to enable individual universities and researchers to plan their research activities and will form the basis for the introduction of a governmental research indicator at a national level.

1 Introduction and background

1.1 Introduction

Over the last 3-4 years there has been a marked shift in the role of CRIS-systems as implemented and used at Danish universities. In fact, if melodramatically inclined, it could be described as a revolution. From a setting where each university did things in their own way with their own systems to the situation today, where universities use common, standardized formats (and even to a large extent use the same CRIS system) – and even agreeing to coordinate research registration and documentation at a very basic level: there where data registration originates.

At the same time, political expectations and demands have changed just as dramatically. Universities' and their researchers are expected to take part in the public debate and to be accessible and understood by the general public, as well as being far more open to cooperation with the private sector. Parallel to these expectations has been the political desire to be able to "objectively" measure the output, performance, penetration and quality of research as practiced by universities. Most universities and researchers have recognised this agenda – and even, though sometimes grudgingly, have accepted it as a political reality.

To accomplish this a dramatic shift in the Danish research information infrastructure was needed. Put in a nutshell we could hear the prompter say: "We need a system which both researchers and research planners can use to administer and document their research activities – that is, their publications, research projects and associated research activities. Furthermore this "system" must enable research documentation in a compatible way across institutions at a national level. "

This new infrastructure is now well on its way to being realised with a resulting dramatic shift in the role of CRIS-systems in Denmark. Prior to this new landscape the main focus of CRIS's in

Denmark had been on the communication of research activities to the surrounding society and less on the documentation of research activities for planners and politicians. The shift to this twofold political climate came at a time when many existing CRIS systems were in the process of being either phased out or their focus was being readjusted, largely due to the new interest in institutional repositories. This also meant that universities did not have to play catch-up: the infrastructure at the systems-level was largely already in place, with only minor adgustments needed to implement the political aims of research evaluation, once they had been formulated.

1.2 Background

The process started in 2004 when 4 universities applied for and got a grant from Denmark's Electronic Research Library (DEFF)[1] to implement an entirely new version of an existing software product PURE, from the Danish software company, Atira A/S[2]. From this initial group of 4, the number has now grown to 8 installations at universities and their faculties. As far as universities are concerned, only The Technical University of Denmark has an implementation of another software product. The PURE software is at the moment in the process of being implemented at other institutions of higher education.

Even though it is a basic requirement today for systems to be able to painlessly exchange standardized data, reality shows that the process is not always that painless! There can be problems at the system level (no matter what the intention) and not least, at the data level. That most universities in Denmark have implemented the same system, undoubtedly alleviates a large part of the "system" problem, while the method of cooperation between the involved universities and how we cooperate with the software company in connection with the underlying metadata formats, alleviates the "data" problem. More on that later. The coordination between universities at various levels in Denmark as it takes place today, is an unprecedented situation and has had a major influence on this new era of research planning and evaluation in Denmark.

1.3 PURE: The system

The PURE platform is a commercial, modular system for the establishment of institutional repositories, which could be termed "i-positories". The term "i-pository" as used here, is a wide term and covers a system which collects, documents and preserves the research output and research activities of a research institution. The description of the PURE system in this article is not intended as an in-depth treatment of all the functions and facilties available in PURE, but is intended to give an overview of the system with special emphasis on how it can be used for research planning and evaluation.

PURE modules can be acquired which cover the registration and documentation needs of a research institution. As regards *content types* there are modules which cover research publications,

---

1 http://www.deff.dk/

2 http://www.atira.dk/

and projects, research activities (such as membership of committees and boards, peer reviewing and editing of journals etc.), press releases and an expert database. With the expert database it is possible to establish an hierarchical set of subject areas and to connect researchers to these areas as a means of research communication to the interested public. Parallel to these core research content types is a module to register student projects and to register publications written while members of another institution, but which are regarded as "external" to a given university and thus not included for example in statistics and other reporting.

In addition to these "content modules" are the *infrastructural modules* whereby the i-pository itself can be maintained, monitored and documented. These modules are used across content modules in the PURE system. Here the backbone is made up of people and their organisations, the central components in a research documentation system. It is essential for a research documentation system to keep close track of the rather dynamic connections between people – people as researchers, people as "employed objects" – and their organisations and to associate these dynamic connections with the various content types. Most PURE implementations are connected to the universities' central administrative systems to ensure that updated personnel/organizational data is available.

As the shift from a research communication system to a research documentation system took place, the university working group which oversees the PURE development needs of universities, realised that an essential component would be a flexible report generator, which would satisfy the needs of politicians and planners. The report generator, which again works across all content types, was one of the first new modules developed after the initial implementation at the universities involved, and is under constant development. It is seen as an essential component which must constantly reflect the documentation needs of research planners.

The report generator can create reports which list the research production (publications, projects , activities etc.) of researchers and their organisations, can create statistical overviews of research production for researchers and their organisations and can report on the bibliometric data available for puiblications (such as citations, impact factor etc.). It is possible to limit by almost all relevant parameters available, such as year, person, organization, publication type, language etc. and to output to a variety of formats, such as HTML, XML, PDF, RTF and CSV (for import into other systems for additional analysis, if needed).

Included in the application to DEFF was the development of other components which were seen as either essential to enable PURE to "fit in" with the existing research information infrastructure or components seen as useful to the members of the consortium. Included as an essential component was the ability to exhibit research data for OAI-PMH harvesting and a Z39.50 module (which was certainly more fashionable a couple of years ago!). One of the members of the consortium, the library of the Faculty of Life Sciences of the University of Copenhagen, had for many years "manually" produced bibliometric analyses of the Faculty's research production, and it was timely that this activity could be implemented as a module coupled to publication content types in PURE.

Another member of the consortium, Roskilde University, had been using DSpace as a fuld text repository, so it was relevant for them to be able to couple DSpace to PURE. As part of a later DEFF project, Atira A/S have developed connectors for both DSpace and Fedora. As yet the Fedora connector has not been used, but is expected to be used when a preservation service for institutional repositories is established in Denmark. This project is at the moment in its initial stages.

One of the most recent additions to PURE (February 2008) is a module available for the Universities of Copenhagen, Aalborg and Aarhus, which enables the importing of PubMed publications in PURE. This is the first of this type of import module, where data from external databases can be pulled into PURE and certainly an area which will be expanded on, as a way of relieving the research registration process at individual universities.

Publication data can be moved out of PURE in various formats: in an internal format, which allows data to be moved to other PURE installations, for example, when a researcher moves from one “PURE university” to another “PURE university”. In the target PURE installation, publications will be tagged as “external” publications and in this way kept separate from all research documentation, while at the same time allowing for a researcher to exhibit their full publication list or maintain a complete CV. Other export formats include, Reference Manager, Bibtex and XML.

Data import/export is an important area that the PURE university working group will be looking at in more detail in the near future. The ambition is to increase the number of external data sources which PURE can stick its trunk into to pull data, so as to lighten the load on the local registration process. Hopefully one day soon, the ideas which are being formulated for a national “data well” will materialise. This “well” is planned as a large national metadata repository which could also be used to feed metadata into PURE.

A central facility in PURE is the means to be able to control a flexible work-flow in the registration proces. This allows universities to set up work flows which support centralized and dentralized registration processes through different roles applied to faculty, department and library staff at varying levels.

To summarize: all in all the main shift that has taken place with the implementation of the PURE system, has been the move away from systems geared for research communication to a system which has been geared for research documentation. The essential basis for research documentation is metadata in which the semantic relationships between elements is explicitly integrated into the model.

To give a bird’s-eye view of the PURE landscape, Figure 1 illustrates the various PURE modules functional add-ons and content types.

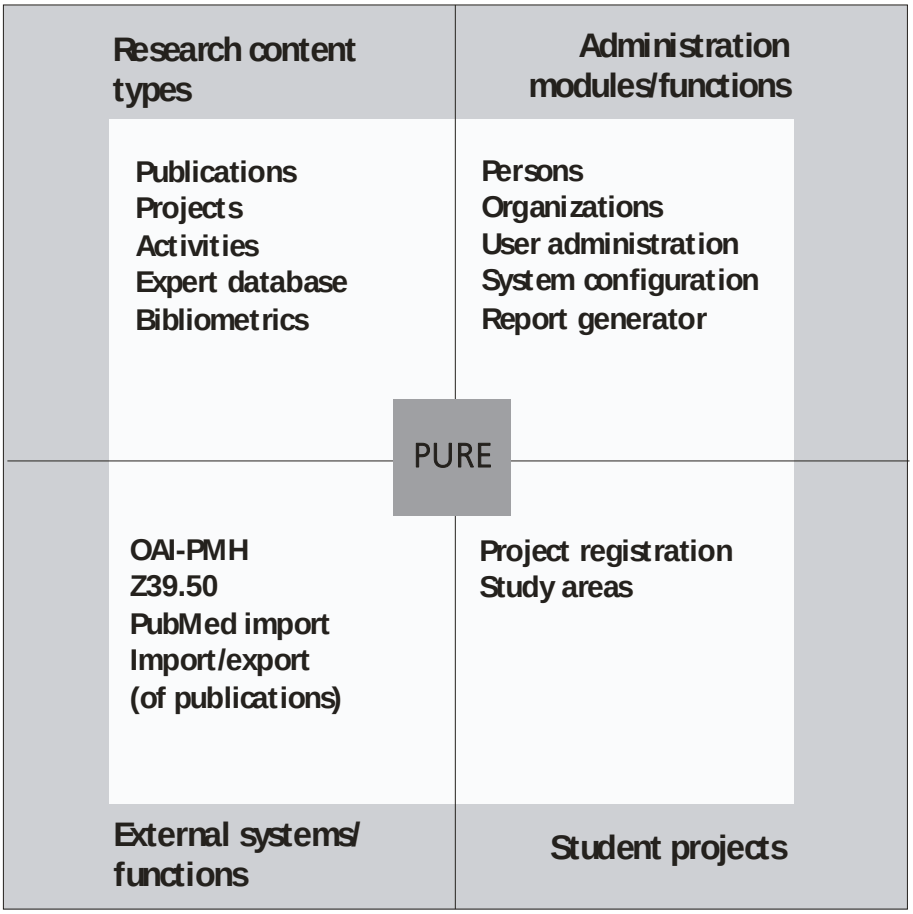

Figure 1: PURE modules, functions and content types

1.4 PURE: The metadata format

At the outset of the initial further development of PURE for the new university consortium in 2004, it was realised that the existing metadata model (in the existing PURE version 1) would require a complete work-over. The consortium established a working group which set to work on rethinking the existing metadata format. The new metadata format was then implemented by Atira A/S and all revisions of the format are watched over by the working group.

The format covers all the content types and other modules in PURE (publications, projects, activities, experts, bibliometrics and student projects). It is not possible to here give an entire overview

of the PURE metadata model, but only a brief overview of how the world is divided into publication types, as a means of categorising the scientific output of researchers, is sufficient.

The top-most division of publication types is fourfold: Research, Communication, Education and Commissioned. At this level the categories are defined mainly according to the nature of the target audience. Research publications are aimed at fellow researchers in which new research results are spread, Communication is aimed at a larger audience in which existing research can be (re)formulated for wider consumption. Education is publications (text books etc.) aimed at students at all levels and Commissioned are publications for or on behalf of governmental and other agencies and authorities.

On the next level is the traditional set of specific and ("traditional") publication types, for example peer reviewed and non-peer reviewed journal articles, books, parts of books, conference articles, reviews, reports etc. All in all can publications be categorised into over one hundred specific types, which are in turn associated to one or more of the four main types.

The PURE metadata format, although at its core rather stable, is constantly watched over and adjusted by the PURE university working group. Its stability is probably largely due to the still rather conservative publication habits and channels of university research. But this can change and the format must be adjusted if and when major changes appear.

1.5 Aggregation at the national level

For many years there has been a (largely) undigested aggregation of research publication (and project) registration at a national level via The Danish Research Database (DDF).[3] The "undigested" version reflected local databases with very little attempt at metadata "streamlining" or even a striving to claim to be complete. This "undigested" version is soon to be replaced by a version based on OAI-PMH harvesting according to a national exchange format for publications (DDF-MXD[4]). By nature this will build largely on the present university PURE installations, which are now certainly more homogenous than previously. This should mark a significant improvement in the quality of data in DDF, which has long been a critical factor in this "national" research database.

It should be stressed, that even though the underlying systems have changed, both locally and at the national level, it is still the responsibility of individual universities to register research publications. This is usually done at the faculty and department level. This factor will always affect the nature and quality of research registration – no matter how nationally coordinated formats and guidelines are.

---

3 http://forskningsbasen.deff.dk/?lang=eng

4 http://tinyurl.com/y5ujn2

2 Benchmarking and data

2.1 The Danish bibliometric indicator

Danish universities have over the last couple of years been engaged in a process of merging after the government devise "size does make a difference". Once independent institutions have become part of other institutions, thus creating larger research institutions. At the same time government has announced, that the funding of research is to be increased, but not as a free luncheon buffet. Universities are expected to compete for funding and funding will also be allocated on the basis of a set of indicators, covering such things as degree of cooperation with private companies, number of Phd. students, patents, educational through-time for students, as well as softer factors such as how an institution communicates with the surrounding society. A central indicator here is a bibliometric indicator, which has been modelled on what is known in Denmark as the "Norwegian model".[5]

In the Danish version of the "Norwegian model", which at this moment is being implemented, publications by researchers at Danish universities will be classified into 5 different publication types: 1. Monographs, 2. Articles in peer reviewed journals, 3. Contributions to monographs/anthologies, 4. Patents and 5. Phd and doctoral theses. Each publication in each of these types, will be given points, according to whether and where they appear on a centrally administered list of "publication channels". For example, will classification into type 2. Aticles in peer reviewed journals, be according to a centrally administered list of journal titles which are "officially" recognized as being peer-reviewed. Articles in journals which are not on this list, i.e. they do not live up to the model's criteria for being "peer-reviewed publication channels", will not receive points in the yearly audit and thus not contribute to an institution's portion of the funding allocated by the bibliometric indicator to the individual universities. Exactly how much research funding will be allocated after the various indicators have been calculated, is at the moment unknown, but it is expected that a ceiling will be applied so as not to have an adverse effect on the year to year planning cycle of universities.

As of writing, how PURE installations at the various "PURE universities" will technically become a part of the infrastructure for this model, has not been decided in detail. It is recognized though, that it will be local PURE installations which will play a central role in the documentation process and will be the starting point for the yearly audit of research publications according to the ranking model laid down. The initial registration of research publications will continue to be the responsibility of the individual universities, but there will be established a number of central, auxiliary databases, which will be utilized in the validation process, to ensure an accurate, objective and conform audit across universities. This is of course an essential requirement for a system which has fund allocation as its end-point.

---

[5] What follows is only meant as a brief description of the coming "bibliometric indicator" in Denmark. What is in focus here is how PURE will be utilized in this system and not the details of the bibliometric indicator itself.

Effort in the initial stages is to establish databases for the officially recognized “publication channels” (both journals and monographs) and how they are to be ranked. The selection of these journals which are considered official “publication channels” has been delegated to a large number of researchers divided into subject groups (68 in all), who have the responsibilty for selecting the relevant journals included in each list and which will be included in a central database of “publication channels”. It is expected that this database will be updated each year according to a set of guidelines given to the various subject groups.

Another central database being planned and essential to the technical infrastructure necessary, will be a database of researchers employed at Danish universities, which will be used to identify them, their period of employment and the institutions which they are a part of. This database will be used to correctly identify author-organization coupling and their involvement in individual publications, in a manner which will, once again, ensure an accurate, objective and conform audit across universities. Both databases will be available to local PURE installations, which will ensure that publications at individual universities are classified in a uniform way and that universities are credited for their authorship. The exact way in which central databases and the various PURE installations will interact is as yet undecided, suffice it to say, that these central databases are expected to be the building blocks for the relevant data in the various local PURE systems.

Turning briefly to the implementation plan: detailed planning of the central databases is expected to be initiated in April-May 2008 and implemented in the following months, together with a number of other systems necessary for the infrastructure required for a centrally available valid set of indicators. These additional systems are for example those needed in the deduplication of publications at the national level and the systems necessary for the calculation of the actual bibliometric indicators and how they will effect yearly funding.

The overall time plan as it is envisaged at the moment at the ministerial level, is that data for publications registered (i.e. published) in 2008 will become the basis for the calculation of the bibliometric indicator, which will result in the allocation of funding for the first time in 2010.

It is expected that the adjustments needed to the present PURE system to be able to fit into the above infrastructure will be minimal, due to several factors. It of course plays a major role that most Danish universities now use the same system, but also of decisive relevance is that universities have accepted a common metadata model and a common development plan for this model. On top of this the PURE system utilizes, and will always utilize, international standards such as OAI-PMH.

But perhaps of paramount importance, as has already been stated, is that the development of PURE is as a system geared to a large extent for *research documentation*, an area central to all future plans for universities’ utilization of PURE. Research documentation is entirely dependent on valid data, data available in open systems and that there is a reliable and flexible means of reporting. The commitment of Danish universities in persuing a common development policy with regards to PURE predates central, ministerial plans for applying research indicators across all

universities, and thus the implementation (both technically and organizationally) is expected to be reasonably trivial. Lucky central planners!

2.2 Data, Data, Data

The shift from research exposure to research documentation, in the registration of publications, projects and activities, has meant that it is necessary to also continually refocus on the pieces of data in i-positories, i.e. is the data stringent and valid and where can we obtain valid data for re-use.

At Danish universities most registration is done locally at the faculty/department level and not always by library staff. Some universities though have installed central validation processes (at least for publications), which does improve the situation somewhat. But experience shows that in the end result there is room for improvement in the validity and consistency of data. This is essential for comparing the research output of universities. The bibliometric indicator described above will help the situation as regards the quality of data for the core research publication types. As described, data will be provided by central auxiliary databases. It is hoped that in time data can also be made available from central databases for the remaining publication types as well as for project and activity registration.

Key data elements wich are notorius for requiring close validation are: the unambiguous identification of researchers (where names are a very poor system of identification!), organisations (where once again names are a poor system), and bibliographic data elements such as publication type, titles etc. Even formal bibliographic identification systems in use (for example ISSN) are used inconsistently. For researchers and their organisations it is hoped that centrally available persistent identifiers will be introduced some time in the near future.

As regards data for reuse, this is an area which will be focused on in the national PURE cooperation. PubMed import has already been mentioned and the PURE national working group are planning to investigate additional sources of metadata. It would even be feasible to reuse data between “PURE universities” which, for example, have common authors. Import of data, while obviously not always ensuring valid data, will ease the burden of registration on individual universities and help consistency.

At the local administrative level there is data available which can be reused by local i-positories. It is expected that the coming central researcher database will be fed by local personnel systems.

2.3 PURE and issues for future development

Apart from the areas already mentioned, there are plans for a DEFF project which will give the PURE metadata format for research projects a workover. Especially in focus will be how to obtain better data, for example not only for Danish organisations but also international organisations which cooperate on research projects with Danish universities. At the moment this data is gener-

ally of rather poor quality: it is not stringent enough to make project data available for in-depth documentation purposes. The above central databases being planned will also be useful for projects and there is data available in central administrative systems, which could be re-used. This is mainly in the areas of project funding and project management. It is unclear at the moment to what extent research project activities will become an area of focus for benchmarking at the national level, but even at the local level this is important data, potentially available to management for day-to-day planning.

One of the activities in this DEFF project will be to produce a "DDF-MXP": a national exchange format for project metadata. If The Danish Research Database is seen as having a future, project data can be harvested from local systems in the same way as publications.

The present version of PURE is 3.x and there are plans to release a PURE version 4 which will be CERIF compliant. This version is expected to be released in early 2009.

A vital area for growth will be the development of PURE at universities into fully-fledged i-positories, that is, containing a comprehensive portion of research production in full text. There have been attempts to improve the situation, but full momentum has as yet not been achieved – neither by researchers themselves or mandated by fund providers.

If individual PURE i-positories are able to develop into comrehensive full text archives, it will be interestiing to see what potential there is for further development to enable for example reference linking and citation analysis (Day 2004, 15-17) or for studies based on, for example, who cooperates on what and to what extent, which could be used in research evaluation. Data mining over time in i-positories could aslo be used to investigate what influence the bibliometric indicator has had on publishing activities, amongst other things.

At the time of writing it is unclear as to what the exact role of The Danish Research Database will be in the future. It is still to be seen how much improved data available for harvesting will effect the use of the aggregated database or if other services will arise. It is still an unanswered question as to what extent a central access point to national research metadata is of interest to the "general public". It is also as yet unclear as to what further research documentation uses there could be for an aggregated database.

3 PURE cooperation and ongoing development

From the beginning the PURE system has been developed and implemented at Danish universities in close cooperation with the vendor Atira A/S. Version 1.0 of PURE was implemented by the University of Aalborg as a solo project, and this can perhaps be regarded as its "beta" implementation. With Version 2.x the other Danish universities came on board and very quickly a close and fruitful cooperation was established, with both the universities invloved and Atira A/S.

As has already been stated a national working group has been established which oversees the ongoing development of the system for Danish university users. This working group, which has representatives from the various PURE installations, has its own budget and it's activities are watched over by a steering committee, where members again come from all the involved universities and which also allocates funds to the working group. It is the working group which has close contact with Atira A/S and through which influence on the system is brought about.

The universities involved see it as vital that the system is dynamic. It is essential that we have a system which is able to be adjusted in response to the rapidly changing environment in which it is placed. To be able to do this it is also necessary to have an organization which can ensure this, as well as a vendor capable of meeting these needs. The universities who make up the working group strive to agree on all added functionality and in close contact with the vendor agree on the specifications and time plan for all adgustments.

This is an unusual approach, at least as far as university library experience with other systems is concerned, for example library OPACs, which typically do not have the same degree of flexibility in their development. It will be interesting to see how this cooperation develops as the customer base grows and if it can be maintained.

4 References

Day, Michael (2004): Institutional repositories and research assessment: A supporting study for the ePrints UK Project. http://eprints-uk.rdn.ac.uk/project/docs/studies/rae/rae-study.pdf

5 Contact Information

Adrian Price
Faculty of Life Sciences Library
University of Copenhagen
Dyrlægevej 10
1870 Frederiksberg C.

Email: ap@life.ku.dk
Tel: +4535332116